

**PCT**WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales BüroINTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

<b>(51) Internationale Patentklassifikation <sup>6</sup> :</b> <b>G10L 5/04</b>	<b>A1</b>	<b>(11) Internationale Veröffentlichungsnummer: WO 99/59134</b> <b>(43) Internationales Veröffentlichungsdatum: 18. November 1999 (18.11.99)</b>
<b>(21) Internationales Aktenzeichen:</b> PCT/DE99/01308 <b>(22) Internationales Anmeldedatum:</b> 3. Mai 1999 (03.05.99)  <b>(30) Prioritätsdaten:</b> 198 21 031.0 11. Mai 1998 (11.05.98) DE  <b>(71) Anmelder (für alle Bestimmungsstaaten ausser US):</b> SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, D-80333 München (DE).  <b>(72) Erfinder; und</b> <b>(75) Erfinder/Anmelder (nur für US):</b> HOLZAPFEL, Martin [DE/DE]; Stabelerstrasse 13, D-80933 München (DE).  <b>(74) Gemeinsamer Vertreter:</b> SIEMENS AKTIENGESELLSCHAFT; Postfach 22 17 34, D-80506 München (DE).		<b>(81) Bestimmungsstaaten:</b> JP, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Veröffentlicht</b> <i>Mit internationalem Recherchenbericht. Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist; Veröffentlichung wird wiederholt falls Änderungen eintreffen.</i>
<b>(54) Title:</b> METHOD AND DEVICE FOR DETERMINING SPECTRAL VOICE CHARACTERISTICS IN A SPOKEN EXPRESSION <b>(54) Bezeichnung:</b> VERFAHREN UND ANORDNUNG ZUR BESTIMMUNG SPEKTRALER SPRACHCHARAKTERISTIKA IN EINER GESPROCHENEN ÄUSSERUNG  <b>(57) Abstract</b> <p>According to the invention, spectral voice characteristics are determined in a natural language expression, whereby the expression is digitized and subjected to a wavelet transformation. The speaker-specific characteristics arise from the different transformation steps of the wavelet transformation. Within the scope of a voice synthesis, these characteristics can be compared with characteristics of other expressions in order to generate a continuously sounding synthetic voice signal for the human ear. Alternatively, the characteristics can also be modified in a targeted manner in order to counteract a perceptive dissonance.</p> <b>(57) Zusammenfassung</b> <p>Es werden spektrale Sprachcharakteristika in einer natürlichsprachlichen Äußerung bestimmt, wobei die Äußerung digitalisiert und einer Wavelet-Transformation unterzogen wird. Aus den unterschiedlichen Transformationsstufen der Wavelet-Transformation gehen die sprecherspezifischen Charakteristika hervor. Diese Charakteristika können im Rahmen einer Sprachsynthese mit Charakteristika anderer Äußerungen verglichen werden, um ein für das menschliche Ohr kontinuierlich klingendes synthetisches Sprachsignal zu erzeugen. Alternativ können die Charakteristika auch gezielt verändert werden, um einer perzeptiven Dissonanz entgegenzuwirken.</p>		

# **LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Beschreibung**Verfahren und Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung**

5

Die Erfindung betrifft ein Verfahren und eine Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung.

- 10 Bei einer konkatenativen Sprachsynthese werden einzelne Laute aus Sprachdatenbanken zusammengesetzt. Um dabei einen für das menschliche Ohr natürlich klingenden Sprachverlauf zu erhalten, sind Diskontinuitäten an den Punkten, wo die Laute zusammengesetzt werden (Konkatenationspunkte) zu vermeiden.
- 15 Die Laute sind dabei insbesondere Phoneme einer Sprache oder eine Zusammensetzung mehrerer Phoneme.

- Eine Wavelet-Transformation ist aus [1] bekannt. Bei der Wavelet-Transformation ist durch ein Wavelet-Filter
- 20 gewährleistet, daß jeweils ein Hochpaßanteil und ein Tiefpaßanteil einer nachfolgenden Transformationsstufe ein Signal einer aktuellen Transformationsstufe vollständig wiederherstellen. Dabei erfolgt von einer Transformationsstufe zur nächsten eine Reduktion der
- 25 Auflösung des Hochpaßanteils bzw. Tiefpaßanteils (engl. Fachbegriff: "Subsampling"). Insbesondere ist durch das Subsampling die Anzahl der Transformationsstufen endlich.

- Die Aufgabe der Erfindung besteht darin, ein Verfahren und
- 30 eine Anordnung zur Bestimmung spektraler Sprachcharakteristika anzugeben, mit deren Hilfe insbesondere eine natürlich wirkende synthetische Sprachausgabe bestimmbar ist.
- 35 Diese Aufgabe wird gemäß den Merkmalen der unabhängigen Patentansprüche gelöst.

2.

Im Rahmen der Erfindung wird ein Verfahren angegeben zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung. Dazu wird die gesprochene Äußerung digitalisiert und einer Wavelet-Transformation unterzogen.

- 5 Anhand unterschiedlicher Transformationsstufen der Wavelet-Transformation werden die sprecherspezifischen Charakteristika ermittelt.

- 10 Dabei ist es insbesondere ein Vorteil, daß bei der Wavelet-Transformation mittels eines Hochpaßfilters und eines Tiefpaßfilters die Äußerung aufgeteilt wird und unterschiedliche Hochpaßanteile bzw. Tiefpaßanteile verschiedener Transformationsstufen sprecherspezifische Charakteristika enthalten.

- 15 Die einzelnen Hochpaßanteile bzw. Tiefpaßanteile verschiedener Transformationsstufen stehen für vorgegebene sprecherspezifische Charakteristika, wobei sowohl Hochpaßanteil als auch Tiefpaßanteil einer jeweiligen Transformationsstufe, also das jeweilige Charakteristikum, getrennt von anderen Charakteristika modifiziert werden kann. Setzt man bei der inversen Wavelet-Transformation aus den jeweiligen Hochpaß- und Tiefpaßanteilen der einzelnen Transformationsstufen wieder das ursprüngliche Signal zusammen, so ist gewährleistet, daß genau das gewünschte Charakteristikum verändert worden ist. Es ist somit möglich bestimmte vorgegebene Eigenarten der Äußerung zu verändern, ohne daß dadurch der Rest der Äußerung beeinflusst wird.

- 30 Eine Ausgestaltung besteht darin, daß vor der Wavelet-Transformation die Äußerung gefenstert, also eine vorgegebene Menge von Abtastwerten ausgeschnitten, und in den Frequenzbereich transformiert wird. Hierzu wird insbesondere eine Fast-Fourier-Transformation (FFT) angewandt.

- 35 Eine weitere Ausgestaltung besteht darin, daß ein Hochpaßanteil einer Transformationsstufe in einen Realteil

und einen Imaginärteil aufgeteilt wird. Der Hochpaßanteil der Wavelet-Transformation entspricht dem Differenzsignal zwischen dem aktuellen Tiefpaßanteil und dem Tiefpaßanteil der vorhergehenden Transformationsstufe.

5

Insbesondere besteht eine Weiterbildung darin, die Zahl der durchzuführenden Transformationsstufen der Wavelet-Transformation dadurch zu bestimmen, daß in der letzten Transformationsstufe, die aus hintereinandergeschalteten Tiefpässen besteht, ein Gleichanteil der Äußerung enthalten ist. Dann ist das Signal als Ganzes darstellbar durch seine Wavelet-Koeffizienten. Dies entspricht der vollständigen Transformation der Information des Signalausschnitts in den Wavelet-Raum.

15

Wird insbesondere nur der jeweilige Tiefpaßanteil weiter transformiert (mittels eines Hochpaß- und eines Tiefpaßfilters), so verbleibt als Hochpaßanteil einer Transformationsstufe das Differenzsignal, wie oben erläutert. Kumuliert man Differenzsignale (Hochpaßanteile) über die Transformationsstufen, erhält man in der letzten Transformationsstufe als kumulierten Hochpaßanteil die Information der gesprochenen Äußerung ohne Gleichanteil.

25

Im Rahmen einer zusätzlichen Weiterbildung sind die sprecherspezifischen Charakteristika identifizierbar als:

a) Grundfrequenz:

30

Die Schwingung des Hochpaßanteils der ersten oder der zweiten Transformationsstufe der Wavelet-Transformation läßt die Grundfrequenz der Äußerung erkennen. Die Grundfrequenz zeigt an, ob der Sprecher ein Mann oder eine Frau ist.

35

b) Form der spektralen Hüllkurve:

Die spektrale Hüllkurve enthält Information über eine Transferfunktion des Vokaltrakts bei der Artikulation.

In einem stimmhaften Bereich wird die spektrale Hüllkurve von den Formanten dominiert. Der Hochpaßanteil einer höheren Transformationsstufe der Wavelet-Transformation enthält diese spektrale Hüllkurve.

c) Spectral Tilt (Rauchigkeit):

Die Rauchigkeit in einer Stimme wird als negative Steigung im Verlauf des vorletzten Tiefpaßanteils sichtbar.

Die sprecherspezifischen Charakteristika a) bis c) sind bei der Sprachsynthese von großer Bedeutung. Wie eingangs erwähnt, bedient man sich bei der konkatenativen Sprachsynthese großer Mengen realgesprochener Äußerungen, aus denen Beispiellaute ausgeschnitten und später zu einem neuen Wort zusammengesetzt werden (synthetisierte Sprache). Dabei sind Diskontinuitäten zwischen zusammengesetzten Lauten von Nachteil, da diese vom menschlichen Ohr als unnatürlich wahrgenommen werden. Um den Diskontinuitäten entgegenzuwirken ist es von Vorteil, direkt die perzeptiv relevanten Größen zu erfassen und ggf. zu vergleiche und/oder einander anzupassen.

Dies kann geschehen durch direkte Manipulation, indem ein Sprachlaut in mindestens einer seiner sprecherspezifischen Charakteristika angepaßt wird, so daß er in dem akustischen Kontext der konkatenativ verknüpften Laute nicht als störend wahrgenommen wird. Auch ist es möglich, die Auswahl eines passenden Lautes daran auszurichten, daß sprecherspezifische Charakteristika von zu verknüpfenden Lauten möglichst gut zueinander passen, z.B. daß den Lauten gleiche oder ähnliche Rauchigkeit zu eigen ist.

Ein Vorteil der Erfindung besteht darin, daß die spektrale Hüllkurve den Artikulationstrakt des Sprechers widerspiegelt und nicht, wie z.B. ein Polstellenmodell, auf Formanten gestützt ist. Weiterhin gehen bei der Wavelet-Transformation

- als nichtparametrischer Darstellung keine Daten verloren, die Äußerung kann stets vollständig rekonstruiert werden. Die aus den einzelnen Transformationsstufen der Wavelet-Transformation hervorgehenden Daten sind linear voneinander unabhängig, können somit getrennt voneinander beeinflußt und später wieder zu der beeinflußten Äußerung - verlustlos - zusammengesetzt werden.
- 10 Weiterhin wird eine Anordnung zur Bestimmung spektraler Sprachcharakteristika angegeben, die eine Prozessoreinheit aufweist, die derart eingerichtet ist, daß eine Äußerung digitalisierbar ist. Daraufhin wird die Äußerung einer Wavelet-Transformation unterzogen und anhand
- 15 unterschiedlicher Transformationsstufen werden sprecherspezifische Charakteristika ermittelt.

- Diese Anordnung ist insbesondere geeignet zur Durchführung des erfindungsgemäßen Verfahrens oder einer seiner vorstehend
- 20 erläuterten Weiterbildungen.

Weiterbildungen der Erfindung ergeben sich auch aus den abhängigen Ansprüchen.

- 25 Ausführungsbeispiele der Erfindung werden nachfolgend anhand der Zeichnung dargestellt und erläutert.

Es zeigen

- 30 Fig.1 eine Wavelet-Funktion;
- Fig.2 eine Wavelet-Funktion, unterteilt nach Realteil und Imaginärteil;
- 35 Fig.3 eine kaskadierte Filterstruktur, die die Transformationsschritte der Wavelet-Transformation darstellt;

Fig.4 Tiefpaßanteile und Hochpaßanteile unterschiedlicher Transformationsstufen;

5 Fig.5 Schritte der konkatativen Sprachsynthese.

Fig.1 zeigt eine Wavelet-Funktion, die bestimmt ist durch

$$10 \quad \psi(f) = c \cdot \left(1 - \left(\frac{f}{\sigma}\right)^2\right) \cdot e^{-\frac{1}{2} \cdot \left(\frac{f}{\sigma}\right)^2} \quad (1),$$

wobei

f die Frequenz,  
 $\sigma$  eine Standardabweichung und  
 15 c eine vorgegebene Normierungskonstante  
 bezeichnen.

Insbesondere ist die Standardabweichung  $\sigma$  bestimmt durch die vorgebbare Stelle des Seitenbandminimums 101 in Fig.1.

20 Fig.2 zeigt eine Wavelet-Funktion mit einem Realteil gemäß Gleichung (1) und einer Hilbert-Transformierten H des Realteils als Imaginärteil. Die komplexe Wavelet-Funktion ergibt sich somit zu

$$25 \quad \Psi(f) = \psi(f) + j \cdot H\{\psi(f)\} \quad (2).$$

Die Konstante c aus Gleichung (1) wird verwendet, um die komplexe Wavelet-Funktion zu normieren:

$$30 \quad \int_{-\infty}^{\infty} \Psi(f) \cdot \bar{\Psi}(f) df = 1 \quad (3),$$

wobei  $\bar{\Psi}$  die konjugiert komplexe Wavelet-Funktion bezeichnet.



Fig.3 zeigt die kaskadierte Anwendung der Wavelet-Transformation. Ein Signal 301 wird sowohl durch einen Hochpaß HP1 302 als auch durch einen Tiefpaß TP1 305 gefiltert. Dabei findet insbesondere ein Subsampling statt, d.h. die Anzahl der abzuspeichernden Werte wird pro Filter reduziert. Eine inverse Wavelet-Transformation gewährleistet, daß aus dem Tiefpaßanteil TP1 305 und dem Hochpaßanteil HP1 304 wieder das ursprüngliche Signal 301 rekonstruierbar ist.

Im Hochpaß HP1 302 wird getrennt nach Realteil Re1 303 und Imaginärteil Im1 304 gefiltert.

Das Signal 310 nach dem Tiefpaßfilter TP1 305 wird erneut sowohl durch einen Hochpaß HP2 306 als auch durch einen Tiefpaß TP2 309 gefiltert. Der Hochpaß HP2 306 umfaßt wieder einen Realteil Re2 307 und einen Imaginärteil Im2 308. Das Signal nach der zweiten Transformationsstufe 311 wird wieder gefiltert, usf.

Geht man von einem (FFT-transformierten) Kurzzeitspektrum mit 256 Werten aus, so werden acht Transformationsschritte durchgeführt (Subsamplingrate: 1/2), bis das Signal aus dem letzten Tiefpaßfilter TP8 dem Gleichanteil entspricht.

In Fig.4 sind verschiedene Transformationsstufen der Wavelet-Transformation, unterteilt nach Tiefpaßanteilen (Figuren 4A, 4C und 4E) und Hochpaßanteilen (Figuren 4B, 4D und 4F) dargestellt.

Aus dem Hochpaßanteil gemäß Fig.4B ist die Grundfrequenz der gesprochenen Äußerung ersichtlich. Neben den Schwankungen in der Amplitude ist deutlich eine überwiegende Periodizität im wavelet-gefilterten Spektrum zu erkennen, die Grundfrequenz des Sprechers. Anhand der Grundfrequenz ist es möglich, vorgegebene Äußerungen bei der Sprachsynthese einander

anzupassen oder passende Äußerungen aus einer Datenbank mit vorgegebene Äußerungen zu bestimmen.

Im Tiefpaßanteil von Fig.4C sind als ausgeprägte Minima und  
5 Maxima die Formanten des Sprachsignalausschnitts (die Länge des Sprachsignalausschnitts entspricht in etwa der doppelten Grundfrequenz) dargestellt. Die Formanten repräsentieren Resonanzfrequenzen im Vokaltrakt des Sprechers. Die deutliche Darstellbarkeit der Formanten ermöglicht eine Anpassung  
10 und/oder Auswahl passender Lautbausteine bei der konkatenativen Sprachsynthese.

Im Tiefpaßanteil der vorletzten Transformationsstufe (bei 256  
Frequenzwerten im Originalsignal: TP7), kann die Rauchigkeit  
15 einer Stimme ermittelt werden. Der Abstieg des Kurvenverlaufs zwischen Maximum Mx und Minimum Mi kennzeichnet den Grad der Rauchigkeit.

Die erwähnten drei sprecherspezifischen Charakteristika sind  
20 somit identifiziert und können für die Sprachsynthese gezielt beeinflußt werden. Dabei ist es insbesondere von Bedeutung, daß bei der inversen Wavelet-Transformation die Manipulation eines einzelnen sprecherspezifischen Charakteristikums nur dieses beeinflußt, die anderen perzeptiv relevanten Größen  
25 bleiben unberührt. Somit kann die Grundfrequenz gezielt verstellt werden, ohne daß dadurch die Rauchigkeit der Stimme beeinflußt wird.

Eine andere Einsatzmöglichkeit besteht in der Auswahl eines  
30 geeigneten Lautabschnitts zur konkatenativen Verknüpfung mit einem anderen Lautabschnitt, wobei beide Lautabschnitte ursprünglich von verschiedenen Sprechern in unterschiedlichen Kontexten aufgenommen wurden. Mit Ermittlung spektraler Sprachcharakteristika kann ein geeigneter zu verknüpfender  
35 Lautabschnitt gefunden werden, da mit den Charakteristika Kriterien bekannt sind, die einen Vergleich von Lautabschnitten untereinander und somit eine Auswahl des

passenden Lautabschnitts automatisch nach bestimmten Vorgaben ermöglichen.

Fig.5 zeigt Schritte einer konkatenativen Sprachsynthese.

- 5 Eine Datenbank wird mit einer vorgegebenen Menge natürlichgesprochener Sprache verschiedener Sprecher erstellt, wobei Lautabschnitte in der natürlichgesprochenen Sprache identifiziert und abgespeichert werden. Es ergeben sich zahlreiche Repräsentanten für die verschiedenen
- 10 Lautabschnitte einer Sprache, auf die die Datenbank zugreifen kann. Die Lautabschnitte sind insbesondere Phoneme einer Sprache oder eine Aneinanderreihung solcher Phoneme. Je kleiner der Lautabschnitt, desto größer sind die Möglichkeiten bei der Zusammensetzung neuer Wörter. So umfaßt
- 15 die deutsche Sprache eine vorgegebene Menge von ca. 40 Phonemen, die zur Synthese nahezu aller Wörter der Sprache ausreichen. Dabei sind unterschiedliche akustische Kontexte zu berücksichtigen, je nachdem, in welchem Wort das jeweilige Phonem auftritt. Nun ist es wichtig, die einzelnen Phoneme in
- 20 den akustischen Kontext derart einzubetten, daß Diskontinuitäten, die vom menschlichen Gehör als unnatürlich und "synthetisch" empfunden werden, vermieden werden. Wie erwähnt stammen die Lautabschnitte von unterschiedlichen Sprechern und weisen somit verschiedene sprecherspezifische
- 25 Charakteristika auf. Um eine möglichst natürlich wirkende Äußerung zu synthetisieren, ist es wichtig, die Diskontinuitäten zu minimieren. Dies kann erfolgen durch Anpassung der identifizierbaren und modifizierbaren sprecherspezifischen Charakteristika oder durch Auswahl
- 30 passender Lautabschnitte aus der Datenbank, wobei ebenfalls die sprecherspezifischen Charakteristika bei der Auswahl ein entscheidendes Hilfsmittel darstellen.

- In Fig.5 sind beispielhaft zwei Laute A 507 und B 508
- 35 dargestellt, die jeweils einzelne Lautabschnitte 505 bzw. 506 aufweisen. Die Laute A 507 und B 508 stammen jeweils aus einer gesprochenen Äußerung, wobei der Laut A 507 deutlich

10

vom Laut B 508 verschieden ist. Eine Trennlinie 509 zeigt an, wo der Laut A 507 mit dem Laut B 508 verknüpft werden soll. Im vorliegenden Fall sollen die ersten drei Lautabschnitte des Lautes A 507 mit den letzten drei Lautabschnitten des Lautes B 508 konkatenativ verknüpft werden.

Es wird entlang der Trennlinie 509 ein zeitliches Dehnen oder Stauchen (vergleiche Pfeil 503) der aufeinanderfolgenden Lautabschnitte durchgeführt, um den diskontinuierlichen Eindruck am Übergang 509 zu vermindern.

Eine Variante besteht in einem abrupten Übergang der entlang der Trennlinie 509 geteilten Laute. Dabei kommt es jedoch zu den erwähnten Diskontinuitäten, die das menschliche Gehör als störend wahrnimmt. Fügt man hingegen einen Laut C zusammen, daß die Lautabschnitte innerhalb eines Übergangsbereichs 501 oder 502 berücksichtigt werden, wobei ein spektrales Abstandsmaß zwischen zwei einander zuordenbaren Lautabschnitten in dem jeweiligen Übergangsbereich 501 oder 502 angepaßt wird (allmählicher Übergang zwischen den Lautabschnitten). Als das Abstandsmaß herangezogen wird insbesondere im Wavelet-Raum der euklidische Abstand zwischen den in diesem Bereich relevanten Koeffizienten.

25

Literaturverzeichnis:

- [1] I. Daubechies: "Ten Lectures on Wavelets", Siam Verlag  
1992, ISBN 0-89871-274-2, Kapitel 5.1, Seiten 129-137.

Patentansprüche

1. Verfahren zur Bestimmung spektraler Sprachcharakteristika  
in einer gesprochenen Äußerung,
  - 5 a) bei dem die Äußerung digitalisiert wird,
  - b) bei dem die digitalisierte Äußerung einer Wavelet-Transformation unterzogen wird,
  - c) bei dem anhand unterschiedlicher Transformationsstufen der Wavelet-Transformation die sprecherspezifischen  
10 Charakteristika bestimmt werden.
2. Verfahren nach Anspruch 1,  
bei dem vor der Wavelet-Transformation eine gefensterte  
15 Transformation der digitalisierten Äußerung in einen Frequenzbereich durchgeführt wird.
3. Verfahren nach Anspruch 2,  
bei dem die Transformation in den Frequenzbereich mittels  
Fast-Fourier-Transformation durchgeführt wird.  
20
4. Verfahren nach einem der vorhergehenden Ansprüche,  
bei dem in jeder Stufe der Wavelet-Transformation ein  
Tiefpaßanteil und ein Hochpaßanteil eines zu  
transformierenden Signals ermittelt werden.  
25
5. Verfahren nach einem der vorhergehenden Ansprüche,  
bei dem ein Hochpaßanteil nach einem Realteil und einem  
Imaginärteil unterteilt wird.
- 30 6. Verfahren nach einem der vorhergehenden Ansprüche,  
bei dem die Wavelet-Transformation mehrere Transformationsstufen umfaßt, wobei die letzte Transformationsstufe einen Gleichanteil der Äußerung in  
einer der Anzahl Transformationsstufen entsprechenden  
35 wiederholten Tiefpaßfilterung liefert.

7. Verfahren nach einem der vorhergehenden Ansprüche,  
bei dem die sprecherspezifischen Charakteristika bestimmt  
sind durch:
- a) eine Grundfrequenz der gesprochenen Äußerung;
  - 5 b) spektrale Hüllkurve;
  - c) einer Rauchigkeit der gesprochenen Äußerung.
8. Verwendung des Verfahrens nach einem der Ansprüche 1 bis  
7 zur Sprachsynthese,
- 10 wobei einzelne sprecherspezifische Charakteristika im  
Hinblick auf eine natürlich klingende Aneinanderreihung  
von Sprachlauten angepaßt werden.
9. Verwendung des Verfahrens nach einem der Ansprüche 1 bis  
15 7 zur Sprachsynthese,  
wobei aus einer vorgegebenen Datenmenge diejenigen  
Sprachlaute anhand einzelner spektraler  
Sprachcharakteristika ausgewählt werden, die eine  
natürlich klingende Aneinanderreihung von Sprachlauten  
20 gewährleisten.
10. Anordnung zur Bestimmung spektraler Sprachcharakteristika  
in einer gesprochenen Äußerung
- 25 mit eineressoreinheit, die derart eingerichtet ist,  
daß folgende Schritte durchführbar sind:
- a) die Äußerung wird digitalisiert;
  - b) die digitalisierte Äußerung wird einer Wavelet-  
Transformation unterzogen;
  - 30 c) anhand unterschiedlicher Transformationsstufen der  
Wavelet-Transformation werden die sprecherspezifischen  
Charakteristika bestimmt.

1/3

FIG 1

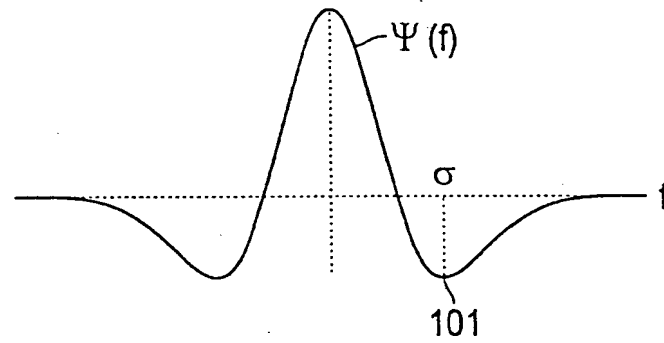


FIG 2

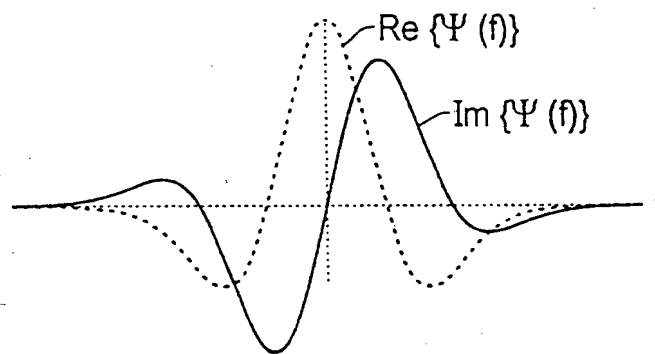
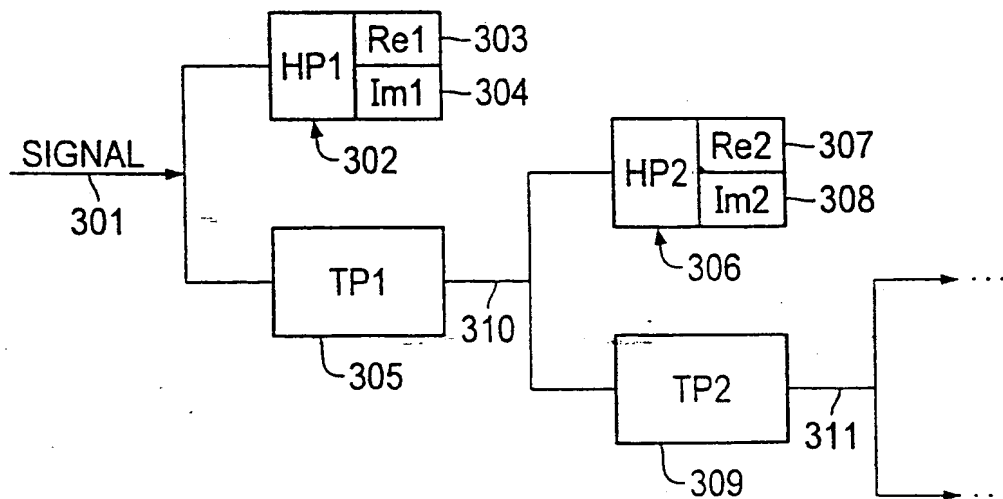


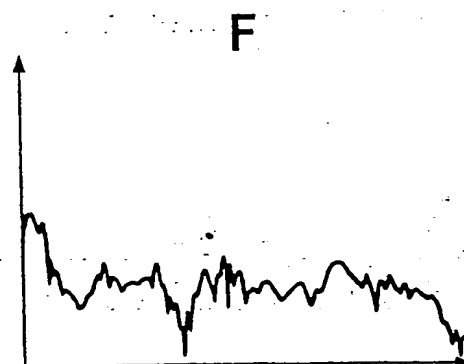
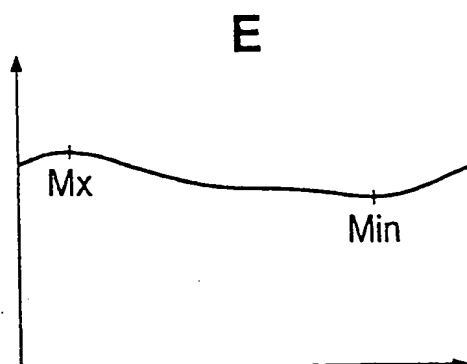
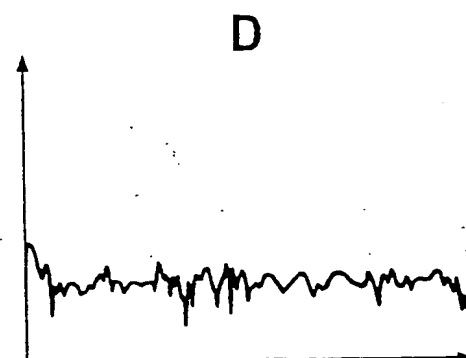
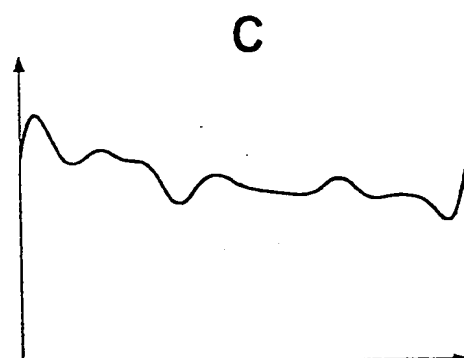
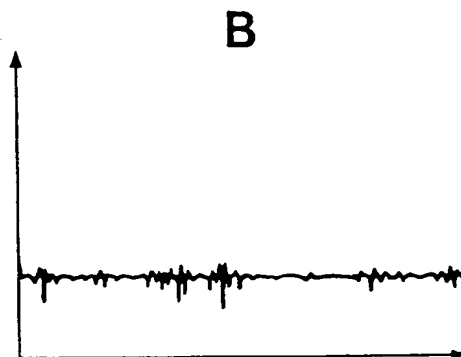
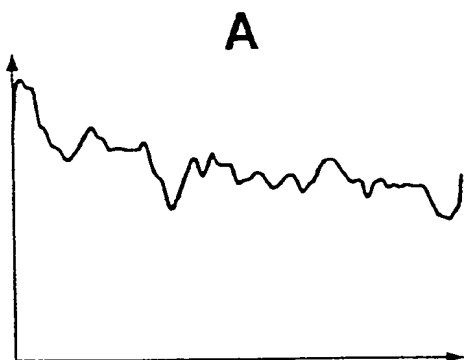
FIG 3





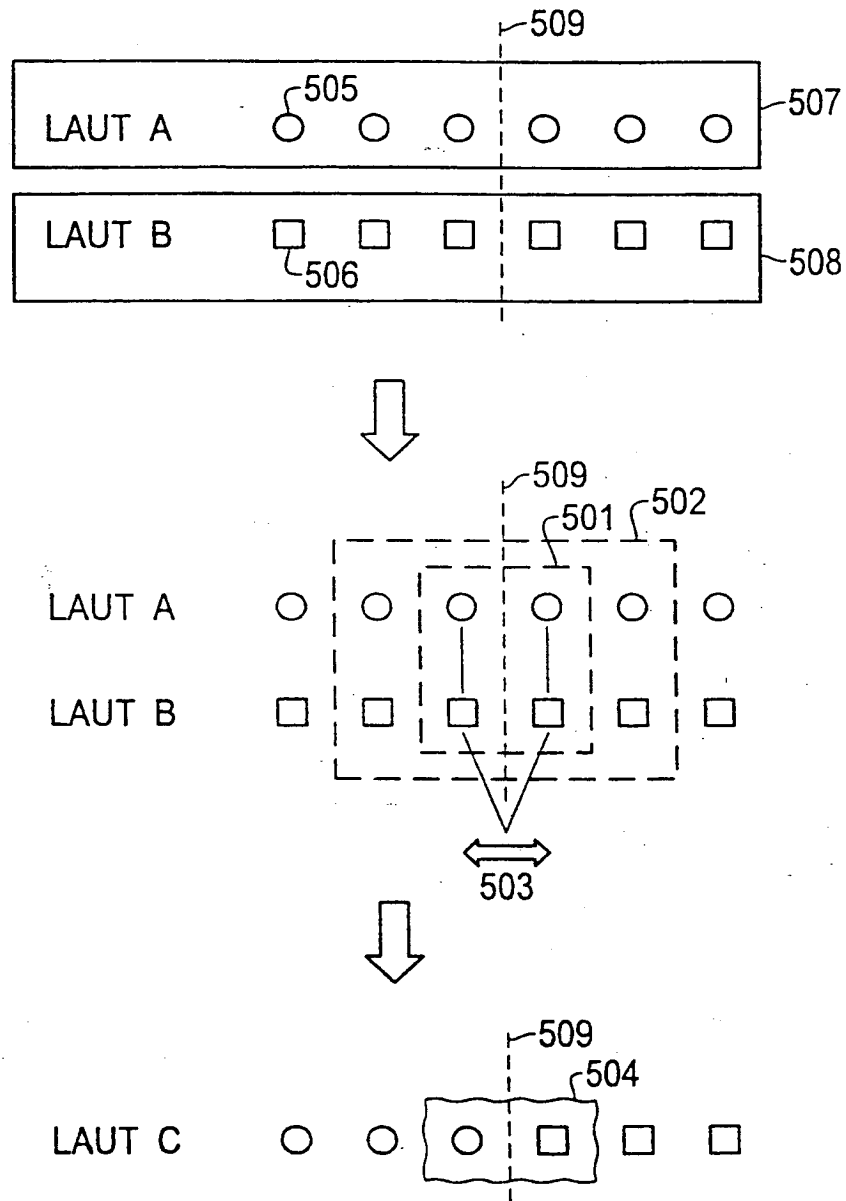
2/3

FIG 4



3/3

FIG 5



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/DE 99/01308

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G10L5/04

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 528 725 A (HUI SIEW K) 18 June 1996 (1996-06-18) column 5, line 9 - line 24; figure 2 column 6, line 57 - column 7, line 47; figure 5	1-4, 6, 7, 10
Y	---	8, 9
Y	EP 0 519 802 A (SEXTANT AVIONIQUE) 23 December 1992 (1992-12-23) page 3, column 3, line 57 - page 4, line 26; figures 3-6	8, 9
A	EP 0 703 565 A (IBM) 27 March 1996 (1996-03-27) abstract page 4, line 5 - page 6, line 34; figures 2, 3	1, 8-10
	--- -/--	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

### \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

5 October 1999

Date of mailing of the international search report

12/10/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Wanzeele, R

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/DE 99/01308

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>EVANGELISTA G: "PITCH-SYNCHRONOUS WAVELET REPRESENTATIONS OF SPEECH AND MUSIC SIGNALS"</p> <p>IEEE TRANSACTIONS ON SIGNAL PROCESSING, vol. 41, no. 12, 1 December 1993 (1993-12-01), pages 3313-3330, XP000426651</p> <p>ISSN: 1053-587X</p> <p>abstract</p> <p>paragraph '00II!; figure 1</p> <p>paragraph '000V!</p> <p>paragraph '0V.A!; figure 7</p> <p>-----</p>	1, 10

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/DE 99/01308

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5528725 A	18-06-1996	GB 2272554 A	18-05-1994
EP 0519802 A	23-12-1992	FR 2678103 A	24-12-1992
		WO 9222890 A	23-12-1992
		JP 6503186 T	07-04-1994
		US 5826232 A	20-10-1998
EP 0703565 A	27-03-1996	JP 8095589 A	12-04-1996
		US 5671330 A	23-09-1997

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/DE 99/01308

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
IPK 6 G10L5/04

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

## B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)  
IPK 6 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

## C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	US 5 528 725 A (HUI SIEW K) 18. Juni 1996 (1996-06-18) Spalte 5, Zeile 9 - Zeile 24; Abbildung 2 Spalte 6, Zeile 57 - Spalte 7, Zeile 47; Abbildung 5	1-4,6,7, 10
Y	---	8,9
Y	EP 0 519-802 A (SEXTANT AVIONIQUE) 23. Dezember 1992 (1992-12-23) Seite 3, Spalte 3, Zeile 57 -Seite 4, Zeile 26; Abbildungen 3-6	8,9
A	EP 0 703 565 A (IBM) 27. März 1996 (1996-03-27) Zusammenfassung Seite 4, Zeile 5 -Seite 6, Zeile 34; Abbildungen 2,3	1,8-10
	---	
	---/---	

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderscher Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderscher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung georacht wird und diese Verbindung für einen Fachmann naheliegend ist

"Z" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

5. Oktober 1999

Absenddatum des internationalen Recherchenberichts

12/10/1999

Name und Postanschrift der Internationalen Recherchenbehörde  
Europäisches Patentamt, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Wanzeele, R

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/DE 99/01308

## C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>EVANGELISTA G: "PITCH-SYNCHRONOUS WAVELET REPRESENTATIONS OF SPEECH AND MUSIC SIGNALS"</p> <p>IEEE TRANSACTIONS ON SIGNAL PROCESSING, Bd. 41, Nr. 12, 1. Dezember 1993 (1993-12-01), Seiten 3313-3330, XP000426651</p> <p>ISSN: 1053-587X</p> <p>Zusammenfassung</p> <p>Absatz '00II!; Abbildung 1</p> <p>Absatz '000V!</p> <p>Absatz '0V.A!; Abbildung 7</p> <p>-----</p>	1,10

# INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/DE 99/01308

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 5528725 A	18-06-1996	GB 2272554 A	18-05-1994
EP 0519802 A	23-12-1992	FR 2678103 A	24-12-1992
		WO 9222890 A	23-12-1992
		JP 6503186 T	07-04-1994
		US 5826232 A	20-10-1998
EP 0703565 A	27-03-1996	JP 8095589 A	12-04-1996
		US 5671330 A	23-09-1997